PATENT ABSTRACTS OF JAPAN

(11)Publication number : 11-112576
(43)Date of publication of application : 23.04.1999

(51)Int.Cl. H04L 12/66
G06F 13/00
H04L 12/46
H04L 12/28
H04L 12/56

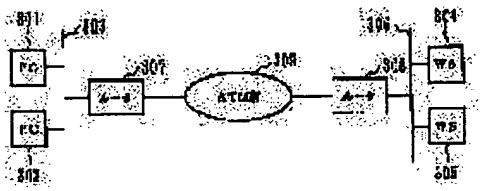
(21)Application number : 09-272832
(22)Date of filing : 06.10.1997

(71)Applicant : HITACHI LTD
HITACHI INFORMATION TECHNOLOGY CO LTD
(72)Inventor : KIYONO TAKASHI
SHIN YOSHIFUMI
SAKOTA HIROYUKI

(54) CONNECTION CONTROL METHOD FOR INTER-NETWORK SYSTEM

(57)Abstract:
PROBLEM TO BE SOLVED: To prevent a channel utilization rate limited by flow control by the transmission control protocol TCP from being deteriorated in a network having a high transmission rate.

SOLUTION: Inter-network systems (routers) 307, 308 monitor TCP connection between adjacent hosts 301, 302, 303, 304, 305 and a host being a communication partner party of any of the adjacent hosts 301, 302, 303, 304, 305 and overcome a delay produced with the pastier host by sending a reception acknowledgment segment to any of the adjacent hosts on behalf of the partner host when detecting deterioration in the channel utilization rate due to the TCP flow control. Furthermore, the routers 307, 308 store data sent from the adjacent hosts in their memories and conduct re-transmission acknowledgment with respect to a re-transmission request from the opposite host by using the data stored in the memories.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(19)日本国特許庁（ＪＰ）　(12)公開特許公報（A）　(11)特許出願公開番号

特開平11－112576

(43)公開日　平成11年(1999)４月23日

(51)Int.Cl.[6]		識別記号	FI		
Ｈ０４Ｌ	12/66		Ｈ０４Ｌ	11/20	Ｂ
Ｇ０６Ｆ	13/00	３５１	Ｇ０６Ｆ	13/00	３５１Ａ
Ｈ０４Ｌ	12/46		Ｈ０４Ｌ	11/00	３１０Ｃ
	12/28			11/20	１０２Ｅ
	12/56				

審査請求　未請求　請求項の数３　ＯＬ　（全 10 頁）

(21)出願番号　特願平9－272832

(22)出願日　平成９年(1997)10月６日

(71)出願人　000005108
株式会社日立製作所
東京都千代田区神田駿河台四丁目６番地

(71)出願人　000153454
株式会社日立インフォメーションテクノロジー
神奈川県秦野市堀山下１番地

(72)発明者　清野　崇
神奈川県海老名市下今泉810番地　株式会社日立製作所オフィスシステム事業部内

(74)代理人　弁理士　磯村　雅俊　（外１名）

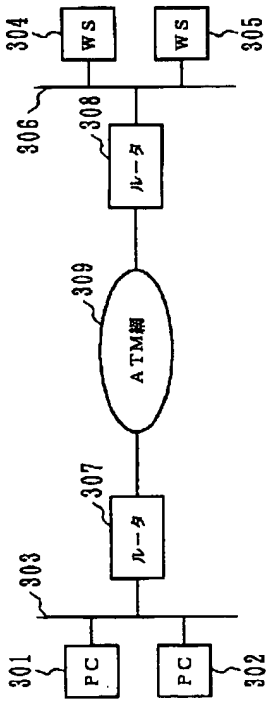
最終頁に続く

(54)【発明の名称】　インターネットワーク装置のコネクション制御方法

(57)【要約】

【課題】高速なネットワークにおいて、ＴＣＰによるフロー制御により制限される回線使用率の低下を防ぐ。

【解決手段】インターネットワーク装置（ルータ）３０７，３０８は、隣接ホスト３０１，３０２，３０４，３０５との通信相手となるホストとの間のＴＣＰコネクションを監視し、ＴＣＰのフロー制御による回線使用率の低下を検出すると、相手ホストに代って隣接ホストに対して受信確認セグメントを送ることにより、相手ホストとの間に生じる遅延時間を克服する。また、ルータ３０７，３０８は、隣接ホストから送信されたデータをメモリ内に保存し、相手ホストの再送要求に対してメモリ内のデータを用いて再送応答する。



1

【特許請求の範囲】

【請求項１】　複数のネットワークを相互に接続し、ＴＣＰ／ＩＰを実装したネットワーク層レイヤ３のＩＰパケットの中継処理を行うインターネットワーク装置のコネクション制御方法において、

前記インターネットワーク装置は、中継受信したデータパケットをＩＰの上位層のＴＣＰレベルに渡し、

該ＴＣＰレベルで該インターネットワーク装置を経由するＴＣＰコネクションの監視を行い、

経由するＴＣＰコネクションを検出した場合、データ送信側ホストが自身と隣接しており、かつＴＣＰのフロー制御による回線使用率が低下しているときには、データ受信側ホストに代って隣接する前記データ送信側ホストに対して受信確認セグメントを送ることを特徴とするインターネットワーク装置のコネクション制御方法。

【請求項２】　請求項１に記載のインターネットワーク装置のコネクション制御方法において、

前記インターネットワーク装置は、ＴＣＰのフロー制御による回線使用率低下を検出するために、ＴＣＰコネクションに関連する回線の帯域幅、インターネットワーク装置自身で計算される回線使用率、ＲＳＶＰによる帯域予約状況、及び隣接ホストとの間のラウンドトリップ遅延時間、ならびに隣接ホストと相手ホストとの間で合意されたウィンドウサイズから検出することを特徴とするインターネットワーク装置のコネクション制御方法。

【請求項３】　請求項１に記載のインターネットワーク装置のコネクション制御方法において、

前記インターネットワーク装置は、隣接するホストからその相手ホストへ送信されるＴＣＰセグメントの信頼性を保証するため、該隣接するホストに受信確認セグメントを送るとともに、該隣接するホストから送信されるデータセグメントをメモリに保存し、

該隣接ホストとの間の順序制御及びエラーチェックを行い、エラー時には該隣接ホストへの再送要求を行い、

前記相手ホストからの再送要求に対しては、保存したメモリ内のデータを用いて該再送要求に応答することを特徴とするインターネットワーク装置のコネクション制御方法。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、複数のネットワークを相互に接続して、ＴＣＰ／ＩＰ（Ｔｒａｎｓｍｉｓｓｉｏｎ　Ｃｏｎｔｒｏｌ　Ｐｒｏｔｏｃｏｌ／Ｉｎｔｅｒｎｅｔ　Ｐｒｏｔｏｃｏｌ）を実装したインターネット装置において、高い回線使用率を実現するコネクション制御方法に関する。

【０００２】

【従来の技術】ある端末の最上位層のアプリケーションレイヤで生成された通信メッセージは、下位層のＴＣＰレイヤおよびインターネットプロトコル（以下、ＩＰと

2

記す）レイヤに伝達され、自端末と相手端末の各ドライバを介して送受信を行い、相手端末のＩＰレイヤから上位層のＴＣＰレイヤに伝達され、さらにアプリケーションレイヤに伝達されることにより、アプリケーション相互間で通信が行われる。この場合、送受信両端末の間をメッセージが経由する他の端末では、ドライバからＩＰレイヤまでしかメッセージは伝達されることはない。従って、ＩＰは信頼性のないデータグラム（ＩＰデータ単位）の配送しか提供していない。この場合のホストーホスト間の信頼性は、ＩＰの上位レイヤに位置するＴＣＰが行う。ＴＣＰは、上位のアプリケーションレイヤに対してコネクションオリエンテッド（接続指向型）なサービスを提供し、通信の信頼性を保証する。ＴＣＰは、通信の信頼性確保のためにチェックサムによるパケットエラーのチェックやシーケンス番号を用いた正当な順序性の確保及び重複の排除を行うのみならず、データ転送の際には受信確認を行い、必要な場合には再送等の手段を用いて信頼性のあるサービスを提供するように努力する。ＴＣＰは、さらにフロー制御により相手が送信することができるデータ量のコントロールを行い、受信側が処理可能な範囲内での通信を行うと同時に、スロースタートアルゴリズムによるネットワーク内で起こる輻輳の回避を行う。

【０００３】ＴＣＰのフロー制御は、スライドする送信ウィンドウに基づいて行われる。このウィンドウ制御を用いることにより、送信側が確認をすることなく複数のデータストリームを連続して送信することが可能となり、その結果、スループットを向上させることができる。ＴＣＰは、本来、エンドーエンド間のプロトコルであって、送信側及び受信側の双方でのみ動作し、中継経路中のインターネットワーク装置等では動作しない。中継経路中のＴＣＰ制御機能に関する従来の技術としては、例えば　パケッティア社のＰａｃｋｅｔＳｈａｐｅｒがある。この装置は、送信側と受信側の間に入って、双方間のＴＣＰ制御パケットを奪い取り、その代りとなるＴＣＰ制御パケットを渡すことにより、帯域幅のコントロールを行い、バースト的に発生するトラフィックの分散を行うことにより、一定のサービスの品質（ＱｏＳ：Ｑｕａｒｉｔｙ　ｏｆ　Ｓｅｒｖｉｃｅ）を確保することが可能となる。また、例えば　特開平８－３３９３５４号公報に記載のネットワーク分散処理システムでは、システム間通信もシステム内通信と同じように、分散共有メモリ経由で行うことを基本とし、システム間で共通の分散共有メモリ空間を共有させる。送信プロセッサモジュールの分散メモリカップラが通信の方向を判断し、各プロトコルに合わせて送信側と受信側の共有メモリの同一アドレスロケーション間でコピーを行うことにより、ＷＡＮ（Ｗｉｄｅ　Ａｒｅａ　Ｎｅｔｗｏｒｋ）を経由したシステム間通信のソフトウェアオーバヘッドを削減し、処理効率を向上させることができる。

【０００４】

【発明が解決しようとする課題】前述の従来技術　では、低速回線を用いた場合において、バースト的に発生するトラフィックの分散を主目的としているため、高速回線を用いたときに起こるＴＣＰのフロー制御によるデータ送信効率の低減を抑制することはできない。ＴＣＰヘッダのウィンドウフィールドは、１６ビット長しかないため、受信確認なしに一度に送ることができるデータのサイズは、約６万５千バイトに制限される。これは少なくとも約６万５千バイトにつき１回は受信確認のやり取りが行われ、その度毎に送信が中断することを意味している。この制限は、ネットワークの帯域幅は狭いときやラウンドトリップ遅延が小さいときには殆んど問題にはならないが、十分な帯域幅が確保された高速なネットワークの場合やラウンドトリップ遅延が大きな場合には、ウィンドウサイズがネックになるため帯域を十分に活用することができない。ＴＣＰの最近の実装においては、ウィンドウスケールオプションを用いてウィンドウサイズの倍率を指定することが可能になっているが、古いＴＣＰの実装や安易なＴＣＰの実装ではこのオプションを用いることはできない。また、ウィンドウスケールオプションを用いた場合においても、最大ウィンドウサイズは約１万６千バイトに制限されるため、将来出現するであろう超高速なネットワークにおいては、やはりウィンドウサイズがネックになりうる。このように、ＴＣＰはフロー制御を行うことにより、信頼性のあるエンドーエンドの通信を確保するが、フロー制御が制限になり、十分に回線の帯域を活用することができないという問題があった。そこで、本発明の目的は、このような従来の課題を解決し、ラウンドトリップ遅延の大きい高速なネットワークにおいても、ウィンドウサイズによる送信ネックを抑制し、回線の使用効率を向上させることが可能なインターネットワーク装置のコネクション制御方法を提供することにある。

【０００５】

【課題を解決するための手段】上記目的を達成するため、本発明によるインターネットワーク装置のコネクション制御方法では、複数のネットワークを相互に接続し、ＴＣＰ／ＩＰを実装したネットワーク層レイヤ３のＩＰパケットの中継処理を行うインターネットワーク装置において、インターネットワーク装置から直接配送されるネットワークに接続される隣接ホストと、その隣接ホストとの通信相手となる相手ホストとの間で行われるＴＣＰコネクションを監視し、ＴＣＰのフロー制御による回線使用率の低下を検出するとともに、前記フロー制御により回線使用率が低下している場合には、相手ホストの代りに隣接ホストに対して受信確認セグメントを送ることにより、相手ホストとの間に生じる遅延時間を克服する。また、前記ＴＣＰのフロー制御による回線使用率の低下を検出する場合に、ＴＣＰコネクションに関連する回線の帯域幅、インターネットワーク装置自身において計算される回線使用率、ＲＳＶＰによる帯域予約状況、及び隣接ホストと相手ホストとの間のラウンドトリップ遅延時間、隣接ホストと相手ホストとの間で合意されたウィンドウサイズから、ＴＣＰのフロー制御による回線使用率低下を検出する。さらに、隣接するホストからその相手ホストへ送信されるＴＣＰセグメントの信頼性を保証するため、隣接するホストから送信されるデータセグメントをメモリに保存し、隣接ホストとの間の順序制御及びエラーチェックを行い、エラー発生時には隣接ホストへの再送要求を行い、相手ホストからの再送要求に対しては、保存したメモリ内のデータを用いて再送要求に応答する。

【０００６】

【発明の実施の形態】以下、本発明の動作原理及びその実施例を、図面により詳細に説明する。

（動作原理）本発明においては、送信ホストに隣接するインターネットワーク装置がＴＣＰコネクションを監視し、各時刻における回線使用率が最適となるようにＴＣＰ受信確認パケットを相手受信ホストの代りに送出することにより、ウィンドウサイズの制限を克服する。通常、インターネットワーク装置の中継処理は、ネットワーク層レイヤ３のみで行われる。インターネットワーク装置は自身が保持しているレイヤ３（ＩＰレベル）の中継処理テーブルの情報を基に受信したデータパケットをどのインターフェースから送出すべきかを判断し、該当するインターフェースに向けて送出する。本発明では、インターネットワーク装置は受信したデータパケットを一旦ＩＰの上位層であるＴＣＰレベルまで押し上げ、インターネットワーク装置を経由するＴＣＰコネクションの監視を行う。ＴＣＰコネクションを検出できない場合には、再びＩＰレベルまでデータパケットが戻され、通常のＩＰレベルでの中継処理を行う。インターネットワーク装置を経由するコネクションを検出した場合、そのインターネットワーク装置は、上記コネクションに関して以下の条件が成立するか否かの検証を行う。

（条件１）データ送信側ホストは自身と隣接しているか否か、（つまり、自身から直接配送のネットワーク上に送信側ホストが接続されているか否か）

（条件２）ＴＣＰコネクションの関連する回線の帯域幅、帯域予約状況、ラウンドトリップ遅延時間、ウィンドウサイズからＴＣＰの最大スループットを計算し、このＴＣＰコネクションに関連した回線の持つ帯域を有効に使用することができないか否か、

【０００７】上記各条件のいずれかでも成立しない場合には、データパケットは再びＩＰレベルまで落されて、通常の中継処理を行う。また、上記各条件が両方とも成立した場合には、本発明におけるインターネットワーク装置は、このコネクション用のバッファを用意し、以後、このコネクションに関連するデータパケットをこの

5

バッファに保存するとともに、再びＩＰレイヤまでデータパケットを戻し、通常のデータ中継処理を行う。データパケットのバッファリングが開始された後に行われる送信側、受信側双方のＴＣＰ制御（データ転送以外のＴＣＰセグメントの送受信）はインターネットワーク装置が肩代りして行う。つまり、送信側からのＴＣＰ制御はインターネットワーク装置がそのＴＣＰ制御に対する応答（さも、受信ホストからの応答であるかのように振舞う）を行い、また受信側からのＴＣＰ制御はインターネットワーク装置がその応答（さも、送信側からの応答であるかのように振舞う）を行う。また、必要な場合には、インターネットワーク装置側からＴＣＰ制御を送信または受信側に対してそれぞれの相手になり代って行う。上記処理を行うことにより、データセグメント送信後の受信確認待ち時間をラウンドトリップ時間からインターネットワーク装置と隣接する送信側ホスト間の２点間の送信時間に短縮することが可能になり、ＴＣＰのフロー制御による回線使用率低下を抑制することが可能になる。

【０００８】データのバッファリングは、ＴＣＰの信頼性を保証するために必要である。例えば、インターネットワーク装置から受信側ホストに到達するまでの間に、送信データが何等かの理由によりデータ化けが起り、受信側ホストがチェックサムエラーを検出した場合を考える。このとき、受信側ホストのＴＣＰは再送要求を送信側ホストに送出する。しかし、この時点で、このデータに対する受信確認セグメントは既にインターネットワーク装置から送信ホストに送られており、送信側ホストのＴＣＰは送信ウィンドウを次へと移しているため、この受信側からの要求に対処することはできない。このような事態に備えて、このインターネットワーク装置ではＴＣＰバッファリングを行い、保存してあるデータを用いて受信側からの再送に対処する。受信ホストからの再送要求に対して自身に保存してあるデータを基にして行うということは、バッファに保存されるデータの信頼性が保証されている必要がある。バッファ内のデータの信頼性を保証するために、このインターネットワーク装置は、バッファをベースとしたＴＣＰ処理（順序制御、エラー処理等）を行う。例えば、送信側から送られたＴＣＰセグメントにおいてチェックサムエラーがあった場合には、それを受けたインターネットワーク装置は再送要求を送信ホストに要求を出し、正しいデータが保存されるように努力する。上記のことを実現するためには、インターネットワーク装置にＴＣＰを実装する必要がある。本来、インターネットワーク装置はネットワークレイヤ３以下でのみ動作し、その上位層まで関知することはないため、ＩＰの上位層であるＴＣＰの実装は必須ではない。しかし、インターネットワーク装置の保守用にｔｅｌｅｎｅｔやｆｔｐのようなアプリケーションが必要になる場合が多く、殆んどのインターネットワーク装置にはＴＣＰが実装されている。そのため、本発明を実現するために既存のプロトコルスタックを改造するだけで、容易に実現できる。また、データの送信側及び受信側双方のプロトコルスタックの変更も必要なく、送信側に隣接するインターネットワーク装置のソフトウェアの変更のみで実現が可能である。

6

【０００９】（実施例）図１は、本発明が適用されるルータ装置（インターネットワーク装置）のモジュール構成図である。本発明のルータ装置は、通信を中継する機能を有し、自身で送信および受信は行わない。ルータ装置は、パケットの中継処理やＴＣＰの制御等のソフトウェア処理を行う主プロセッサ（ＣＰＵ）１０１と、動作するソフトウェアをロードする目的の他に各種通信制御のバッファとして用いられる主記憶装置（メモリ）１０２と、このルータ装置上で動作させる制御ソフトウェアや構成定義情報を格納するためのディスク装置１０３と、複数の各種通信回線を制御するための回線制御機構１０５，１０６とから構成され、各装置は装置内バス１０４により接続されている。ここで構成定義情報は、ルーティングプロトコルに関する情報やアドレス情報等が登録されている。ルータ装置は、電源が投入されると、ディスク装置１０３から制御ソフトウェアが主記憶装置１０２にロードされ、このソフトウェアが主プロセッサ１０１により処理される。制御ソフトウェアは、ディスク装置１０３内に格納されている構成定義情報を読み込み、各種回線制御装置を動作させるとともに構成定義情報に従ったパケット中継処理を行う。

【００１０】図２は、図１のルータ装置上で動作するソフトウェアの構成図である。通信の制御を行うソフトウェアは、図２に示す階層構造を形成している。最下層レイヤには、各種の異なる通信回線制御装置を制御するためのドライバ部２０１があり、その上位層には、ＩＰデータグラムの中継処理、その他のＩＰ処理を行うＩＰ処理部２０２と、ＣＳＭＡ／ＣＤ（Ｃａｒｒｉｅｒ Ｓｅｎｓｅ Ｍｕｌｔｉｐｌｅ Ａｃｃｅｓｓ ｗｉｔｈ Ｃｏｌｌｉｓｉｏｎ Ｄｅｔｅｃｔｉｏｎ）方式のＬＡＮ（Ｌｏｃａｌ Ａｒｅａ Ｎｅｔｗｏｒｋ）において、ＩＰアドレスとハードウェアアドレスとの間の対応付けを行うＡＲＰ（Ａｄｄｒｅｓｓ Ｒｅｓｏｌｕｔｉｏｎ Ｐｒｏｔｏｃｏｌ）処理部２０３と、システム間の種々の調停のため制御コードをやり取りするためのＩＣＭＰ（ＩｎｔｅｒｎｅｔＣｏｎｔｒｏｌ Ｍｅｓｓａｇｅ Ｐｒｏｔｏｃｏｌ）処理部２０４と、実時間処理の必要な通信に対して帯域を予約するためのプロトコルであるＲＳＶＰ（Ｒｅｓｏｕｒｃｅ Ｒｅｓｅｒｖａｔｉｏｎ Ｐｒｏｔｏｃｏｌ）を処理するためのＲＳＶＰ処理部２０５とが設けられる。さらにその上位層には、それぞれＴＣＰ（Ｔｒａｎｓｍｉｓｓｉｏｎ Ｃｏｎｔｒｏｌ Ｐｒｏｔｏｃｏｌ）、ＵＤＰ（Ｕｓｅｒ Ｄａｔａｇｒａｍ Ｐｒｏｔｏｃｏｌ）を処理するためのＴ

ＣＰ処理部２０６およびＵＤＰ処理部２０７と、このルータ装置間を経由するＴＣＰコネクションを監視し、ウィンドウサイズによる通信ネックとなる場合に代替受信確認セグメントを送出して帯域幅の有効活用を行うコネクション制御部２０８とが設けられる。なお、ＴＣＰはコネクションを張って接続するのに対して、ＵＤＰはコネクションを張らずに制御セグメントを送出するのみである。さらに、最上位層には、他のルータ装置との経路情報のやり取りを行い、ＩＰ処理部２０２が参照する経路情報テーブルを更新するルーティング処理部２０９と、アプリケーションプログラムであるｆｔｐ（ｆｉｌｅ　ｔｒａｎｓｆｅｒｐｒｏｔｏｃｏｌ）と、ｔｅｌｎｅｔとが設けられる。

【００１１】図３は、本発明のルータ装置を用いたネットワーク構成例を示す図である。図３において、３０７及び３０８が、上記構成を持つ本発明のルータ装置である。２台のパーソナルコンピュータ（以下、ＰＣ）３０１，３０２及びルータ装置３０７は、１００Ｍビット／秒の帯域幅を持つＬＡＮ３０３により接続され、また同様に２台のワークステーション（以下、ＷＳ）３０４，３０５とルータ装置３０８も１００Ｍビット／秒のＬＡＮ３０６により接続される。２台のルータ装置３０７と３０８の間は、ＡＴＭスイッチ群により構成されるＡＴＭネットワーク網３０９に接続され、１５５Ｍビット／秒の帯域が確保されるが、ルータ装置間の距離は非常に離れており（例えば、東京ーサンノゼ間）、ネットワークの遅延は非常に大きいものとする。ＰＣ３０２のＴＣＰはウィンドウスケールオプションを実装しておらず、そのため最大ウィンドウサイズは６５５３６バイトであるとする。いま、図３のＰＣ３０１とＷＳ３０４の間でビデオ会議アプリケーションを用いた通信を行っているものとする。このアプリケーションは、ＲＳＶＰを用いて３０Ｍバイト／秒の帯域予約を行っているものとする。このビデオ会議アプリケーションはエンドーエンド間の通信にＵＤＰを用いているため、ＴＣＰのコネクションは存在しないものとする。この時、ＰＣ３０２とＷＳ３０５の間において、ＴＣＰの上位アプリケーションであるｆｔｐを用いたファイル転送を行う場合を考える。

【００１２】図４は、図３におけるＷＳ３０５からＰＣ３０２へデータ転送する際の通常の通信シーケンスチャートである。ここでは、通常のＴＣＰを用いてデータ転送している場合を示している。先ず、ｐｈａｓｅ１では、ＷＳ３０５からＰＣ３０２に対してシーケンス２６３０００～３２８５３６のデータセグメントの送信を行い、ｐｈａｓｅ２においてＰＣ３０２からＷＳ３０５に対してｐｈａｓｅ１で受信したデータの受信確認及び次に行われるｐｈａｓｅ３のウィンドウサイズ６５５３６バイトの指定を行っている。同様に、ｐｈａｓｅ３では、次の６５５３６バイトのセグメントの送信を行い、ｐｈａｓｅ４でその確認を行っている。図４の右側に示したＴｔ及びＴｄはそれぞれ６５５３６バイトを送信するのに要した時間と、ネットワークのラウンドトリップ遅延時間を示している。このように、ＰＣ３０２とＷＳ３０５の間には有効な帯域幅が７０Ｍビット／秒存在する。すなわち、ＰＣ３０１とＷＳ３０４との間では、ＬＡＮの１００Ｍビット／秒の帯域予約をＲＳＶＰを用いて行っているので、ＰＣ３０２とＷＳ３０５の間の通信にはその帯域を差し引いた残りの帯域のみで行う必要があり、実際に使用できるのは１００Ｍビット／秒から３０Ｍビット／秒を引いた値となる。さらに、Ｔｄの遅延が存在するため、７０Ｍビット／秒のうちの（Ｔｔ／Ｔｄ）×１００％しかデータ転送には用いられない。この例のように、高速なネットワークで、かつ遅延時間が大きいときほど、ウィンドウサイズの制限がネットワークの回線使用率に重大な影響を及ぼす結果になる。

【００１３】図５は、本発明の一実施例を示す代理受信確認を行う場合の通信シーケンスチャートである。前述のように、高速なネットワークで、かつ遅延時間が大きい場合には、ウィンドウサイズの制限がネットワークの回線使用率に大きな影響を及ぼす。この問題を克服するため、本発明においては、送信側に隣接するルータ装置３０８がＴＣＰの受信確認セグメントをＰＣ３０２の代りに発行し、ＷＳ３０５に送ることにより遅延時間による回線使用率の低下を抑制する。図５において、ルータ装置３０８はｗｓ３０５とＰＣ３０２とのＴＣＰコネクションを監視し、回線の帯域幅を効率的に使用するようにルータ装置３０８がＰＣ３０２の代りにＷＳ３０５に対して受信確認セグメントを発行すると同時に、ＰＣ３０２からＷＳ３０５に対して送信されてくる受信確認セグメントを削除する。図５に示すように、ルータ装置３０８がＰＣ３０２の代りに受信確認をＷＳ３０５に送ることにより、ルータ装置３０８とＰＣ３０２との間の送受信遅延を取り除くことが可能となるので、高いスループットを得ることができる。

【００１４】本発明のルータ装置は、ＴＣＰのウィンドウサイズの制限により回線の持つ帯域幅を有効に使用することができない場合においてのみ代理で受信確認セグメントの発行を行い、その他の場合にはこの処理を行わない。ルータ装置３０８は、１００Ｍビット／秒の帯域幅を持つＬＡＮインタフェースと１５５Ｍビット／秒の帯域幅を持つＡＴＭインタフェースの２つのインタフェースを装備しており、ここではそれぞれｌａｎ、ａｔｍと表現することにする。ｌａｎ及びａｔｍインタフェースの有効な帯域幅Ｂ１及びＢ２は、それぞれ次式（１）、次式（２）で表わされる。

$$B1 = Batm \times (Ra / 100) - Br \quad \cdots\cdots\cdots\cdots (1)$$

$$B2 = Blan \times (Rl / 100) - Br \quad \cdots\cdots\cdots\cdots (2)$$

ここで、Ｂａｔｍ及びＢｌａｎはそれぞれのインタフェースが持つ帯域幅、Ｒａ及びＲｌはそれぞれ ＡＴＭ及びＬＡＮインタフェースの回線使用率、またＢｒはルータ装置において割り当てられたＲＳＶＰによる予約帯域を

Ｂｂ＝ｍｉｎ（Ｂ１，Ｂ２）　・・・・・・・・・・・・・・・・・・（３）

【００１５】一方、ＷＳ３０５とＰＣ３０２の間のラウンドトリップ遅延時間をＴ、またＷＳ３０５とＰＣ３０２の間で合意したウィンドウサイズをＷとすると、このＴＣＰコネクションでの最大スループットＢｗは次式（４）で表わされる。

Ｂｗ＝Ｗ／Ｔ　・・・・・・・・・（４）

ここで、ウィンドウサイズＷはウィンドウスケールオプションを考慮した値であっても差し支えない。もし、前式（４）のＢｗより前式（３）のＢｂの方が大きい場合には、ＴＣＰのウィンドウサイズによる回線使用制限が発生していることになる。一方、Ｂｂの方がＢｗより大きい場合には、ＴＣＰのウィンドウサイズによる回線使用制限よりむしろ有効帯域幅がネックになっていると言える。そのためにルータ装置３０８では、ＷＳ３０５とＰＣ３０２との間のＴＣＰコネクションを監視し、ルータ装置が持っているその時点の回線使用率及び帯域予約状況から次式（５）を算出し、式（５）が真になる場合にのみ代理受信確認応答処理を行う。

（Ｂｂ＞Ｂｗ）　・・・・・・・・（５）

これまでに述べた隣接するルータ装置による代理受信確認による方法は、ＷＳ３０５からＰＣ３０２へのデータ送信で全てエラーがなく届いたときのみ有効に機能する。しかし、ルータ装置３０８からＰＣ３０２へのデータ送信過程で新一タにエラーが発生した場合には、回復する方法がない。すなわち、ＷＳ３０５は、ルータ装置３０８から受信確認セグメントを受信した後に、ＰＣ３０２からの再送要求が届いても、混乱してしまうと予想される。ここでＷＳ３０５は、先に受信した受信確認セグメントをＰＣ３０２から来たものであると思っているからである。このため、本発明においては、ルータ装置３０８がデータのバッファリングを行い、ＰＣ３０２からの再送要求に対応できるようにする必要がある。

【００１６】図６は、本発明の他の実施例を示す再送要求に対処したシーケンスチャートである。図６では、ＷＳ３０５からシーケンス番号２６３００から３２８５３６までのＴＣＰセグメントが送信され、ルータ装置３０８からＰＣ３０２までの間にエラーがあったセグメントのみの再送要求を行っている。しかし、このデータに対する受信確認セグメントの送出は、ルータ装置３０８からＷＳ３０５に対して既に行われているため、ＷＳ３０５に対して再送要求を送ることはＴＣＰの動作に矛盾が起ってしまう。このような問題を解決するために、本発明においては、ルータ装置３０８はＷＳ３０５から送られる全てのデータをルータ装置３０８上のメモリ１０２に保存し、ＰＣ３０２からの再送要求時にはこの保存された示す。ＷＳ３０５からＰＣ３０２への通信時には、ｌａｎ，ａｔｍ双方のインタフェースを使用するため、この通信に有効な帯域幅Ｂｂは次式（３）で表現できる。

データを参照して再送要求に応答する。このデータは、ＰＣ３０２からのこのデータに対する受信確認セグメントが送られてくるまでメモリ１０２内に保存される。

【００１７】一旦、ＰＣ３０２から再送要求が届くと、ルータ装置３０８は自身に貯えられたＷＳ３０５からのデータを用いてそれまでに行われた通信の復旧を行い、復旧が終了するまでＷＳ３０５には受信確認セグメントの送出は行わない。復旧処理が完了した時点で、ルータ装置３０８はＷＳ３０５に対して受信確認セグメントの送出を行い、通信を再開させる。このように、ルータ装置３０８はＷＳ３０５とＰＣ３０２間のＴＣＰコネクションにおいて行われるデータを自身のメモリ内に保存することにより、再送要求に対応させている。しかし、これはルータ装置３０８内のメモリに保存されるデータにエラーがないこと、そして順序が正しいことを前提としている。このことを保証するために、ルータ装置３０８は通常のＴＣＰによるデータ転送と同様のことをＷＳ３０５との間で行い、保存するデータの妥当性を保証する。例えば、ＷＳ３０５からＰＣ３０２宛に送信されたデータセグメントにエラーを見つけた場合には、ルータ装置３０８はそのままＰＣ３０２に対してデータを渡さずに、ＷＳ３０５に対して再送要求を行い、エラーの修復を行う。また、順序通りに届かなかったデータセグメントに対しては、自身のメモリ内においてデータ順序の再構成を行う。

【００１８】図６では、ＷＳ３０５からルータ装置３０８にはデータ１～５を正常に転送したので、ルータ装置３０８はメモリにデータ１～５をバッファリングするとともに、ＰＣ３０２に対して転送するとともに、ルータ装置３０８からＷＳ３０５に対して代理受信確認を行う。しかし、ルータ装置３０８とＰＣ３０２との通信経路中にエラーが発生したため、ＰＣ３０２からデータ４について再送要求が返送された。その時点では、ＷＳ３０５からデータ６～１０の送信が終了し、ルータ装置３０８から代理受信確認が返送されている。そして、ＷＳ３０５からはデータ１１～１５がルータ装置３０８に対して転送されている。ルータ装置３０８は、このデータ１１～１５をメモリに格納する。ルータ装置３０８は、その時点で以降の代理受信確認を中止し、ＰＣ３０２に対してメモリに保存していたデータ４をＰＣ３０２に再送する。ＰＣ３０２からＡｃｋ６が返送されると、ルータ装置３０８は次のデータ６～１０までをメモリから取り出して送信する。ＰＣ３０２からＡｃｋ１１が返送されると、ルータ装置３０８はメモリに保存しているデー

タ11～15をPC302に転送する。PC302からAck16が返送されると、ルータ装置308はメモリにデータが残っていないため、その時点でWS305に対して代理受信確認を行う（Ack16）。

【0019】図7は、本発明の一実施例を示すルータ装置のパケット中継処理のフローチャートである。このフローは、図1に示したルータ装置のLAN回線制御機構105及びATM回線制御機構106からCPU101へのハードウェア割り込み処理により起動される（START）。図3におけるLAN306からのデータ受信により、図2に示すルータ装置308のLANドライバ部201がLANフレームの受信処理を行い（ステップ701）、そのデータをIP処理部202に渡す。IPデータグラムを受けたIP処理部202はIPの受信処理を行い（ステップ702）、中継処理を行う前に上位レイヤであるTCP制御部206に引き渡す。このTCP処理部206が本発明におけるTCPコネクション制御を行う部分である。TCP制御部206はコネクションの監視を行い、TCPコネクション制御が有効な場合においてのみバッファリング、TCPの肩代り処理を行う（ステップ703）。バッファリングや肩代り処理を行った後、IP処理部202に渡し、IP中継処理を行い（ステップ704）、さらにLANドライバ部201に引き渡して、送信処理を行う（ステップ705）。一方、TCPコネクション制御が必要ない場合には、再びIP処理部202にIPデータグラムが戻され、ルーティングテーブルの情報をもとにIP中継処理を行い（ステップ704）、適切なインタフェースからIPデータグラムが送出される（ステップ705）。

【0020】図8は、図7におけるTCP制御部の詳細フローチャートである。TCP制御部206では、先ずTCPのデータであるか否かを判断し（ステップ801）、隣接ホストからの送信であるか否かのチェックを行う（ステップ802）。どちらかが該当しない場合には、TCPコネクション制御を行わずにそのままIPデータグラムの中継処理に戻す（ステップ811）。どちらにも該当する場合には、既にコネクション制御を行っているか否かを判別し（ステップ803）、コネクション開始前のものであれば、コネクション確立用のTCPセグメントであるか否かの検査を行う（ステップ804）。その結果、コネクション確立用のTCP制御セグメントである場合には、前式（5）つまり（Bb＞Bw）の判定を行い（ステップ805）、真の場合にはコネクションの登録及び監視を開始する（ステップ806）。一方、既にコネクションの登録が行われていた場合には（ステップ803）、データのバッファリングを行った後（ステップ807）、このバッファをベースとした代替えTCP処理を行う（ステップ808）。このバッファベースのTCP処理では、これまでに述べた代理TCP受信セグメントの送出やバッファ内の順序制御、再送要求、再送要求に対する代理応答等が行われる。TCP処理が終了した後、必要なデータをIP送信処理に渡し（ステップ809）、不要になったバッファ内のデータの削除を行い、バッファを開放して処理を終了する（ステップ810）。

【0021】

【発明の効果】以上説明したように、本発明によれば、送受信を行うホストのプロトコルスタックを変更することなく、高速なネットワークにおいてTCPによるフロー制御により制限される回線使用率の低下を防ぐことができ、高いスループットを得ることができる

【図面の簡単な説明】

【図1】本発明が適用されるルータ装置のハードウェア構成図である。

【図2】図1におけるルータ装置のソフトウェア構成図である。

【図3】本発明が適用されるネットワーク構成例を示す図である。

【図4】本発明を使用しない通常のTCP動作のシーケンスチャートである。

【図5】本発明の一実施例を示す代理受信確認動作のシーケンスチャートである。

【図6】本発明の他の実施例を示す再送発生時の動作シーケンスチャートである。

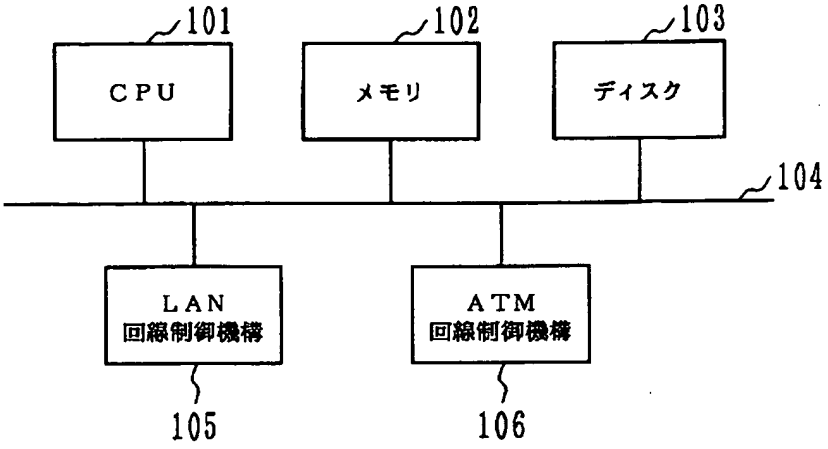
【図7】本発明のルータ装置における中継処理のフローチャートである。

【図8】図7におけるTCP制御部の処理フローチャートである。

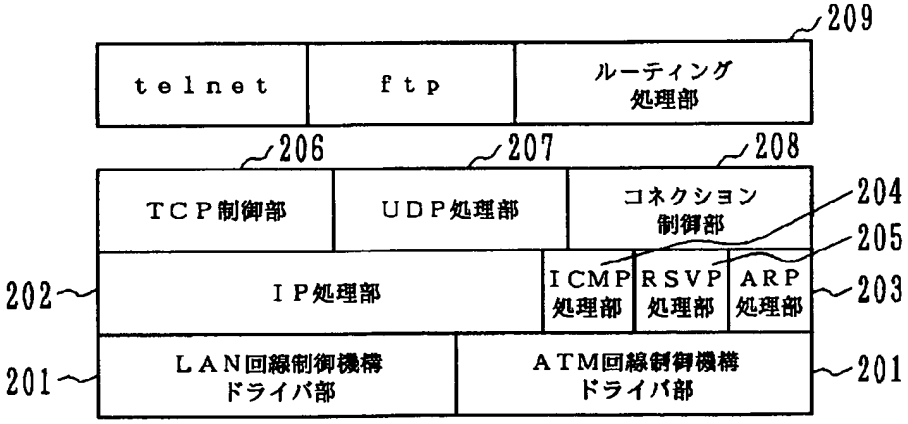
【符号の説明】

101…CPU、102…メモリ、103…ディスク、104…内部バス、105…LAN回線制御機構、106…ATM回線制御機構、201…LAN回線制御機構ドライバ部、ATM回線制御機構ドライバ部、202…IP処理部、203…ARP処理部、204…ICMP処理部、205…RSVP処理部、206…TCP処理部、207…UDP処理部、208…コネクション制御部、209…ルーティング処理部、301，302…PC、304，305…WS、307，308…ルータ装置、309…ATM網、303，306…LAN。

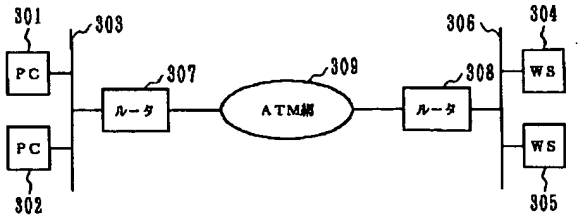
【図1】


101
CPU
102
メモリ
103
ディスク
104
LAN
回線制御機構
ATM
回線制御機構
105
106

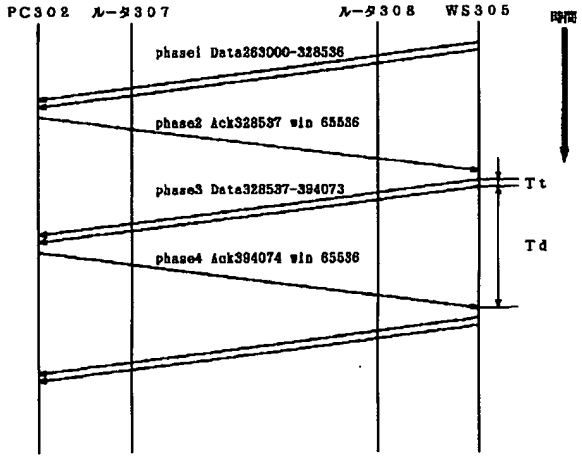
【図2】


209
telnet
ftp
ルーティング
処理部
206
207
208
TCP制御部
UDP処理部
コネクション
制御部
204
205
202
IP処理部
ICMP
処理部
RSVP
処理部
ARP
処理部
203
201
LAN回線制御機構
ドライバ部
ATM回線制御機構
ドライバ部
201

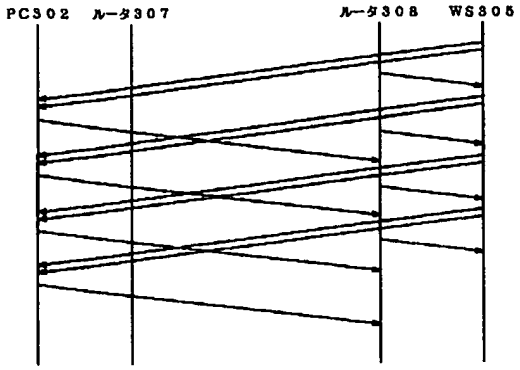
【図3】


301
303
306
304
PC
307
309
308
WS
PC
ルータ
ATM網
ルータ
WS
302
305

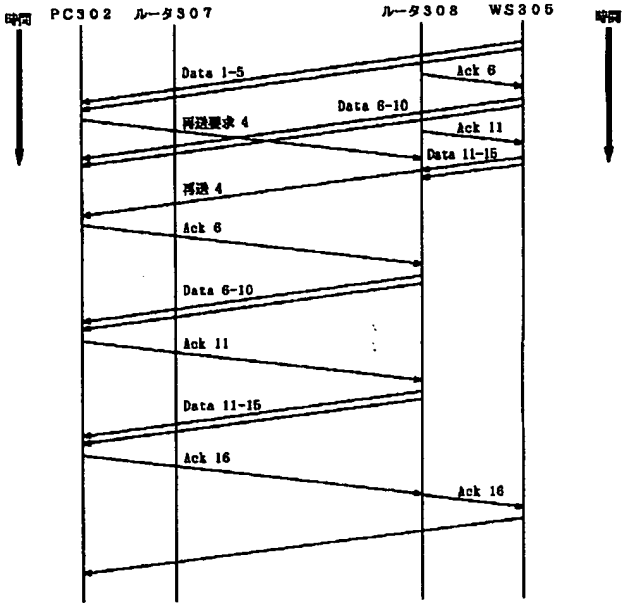
【図4】


PC302
ルータ307
ルータ308
WS305
時間
phase1 Data263000-328536
phase2 Ack328537 win 65536
phase3 Data328537-394073
phase4 Ack394074 win 65536
Tt
Td

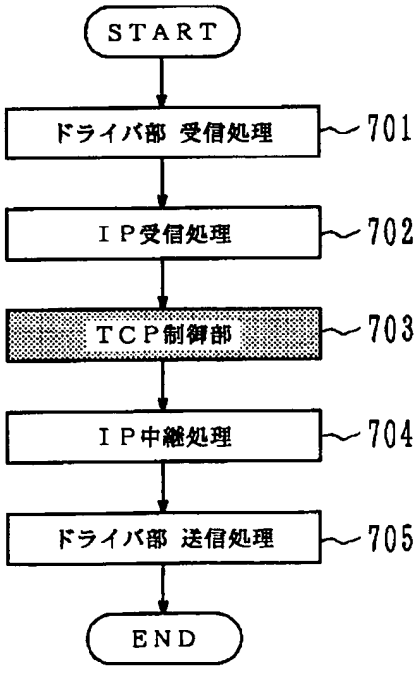
【図5】


PC302
ルータ307
ルータ308
WS305

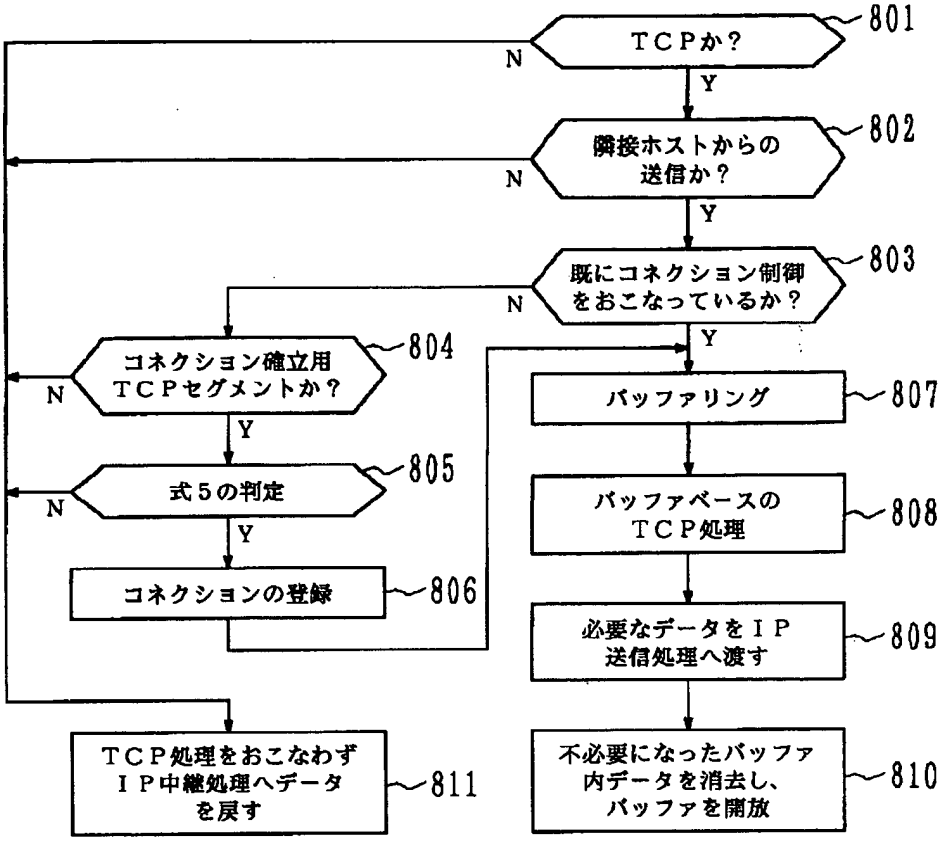
【図6】


時間
PC302
ルータ307
ルータ308
WS305
時間
Data 1-5
Ack 6
Data 6-10
再送要求 4
Ack 11
Data 11-15
再送 4
Ack 6
Data 6-10
Ack 11
Data 11-15
Ack 16
Ack 16

【図7】


START
ドライバ部 受信処理
701
ＩＰ受信処理
702
ＴＣＰ制御部
703
ＩＰ中継処理
704
ドライバ部 送信処理
705
END

【図８】



フロントページの続き

(72)発明者　新　善文
神奈川県海老名市下今泉810番地　株式会社日立製作所オフィスシステム事業部内

(72)発明者　迫田　博幸
神奈川県秦野市堀山下１番地　株式会社日立インフォメーションテクノロジー内